Limitations to Modeling and Analysis Techniques

"After the Crashes are Counted", TRB Workshop

By Simon Washington
University of Arizona, Tucson AZ

Soon to be: Visiting Professor,
Urban Engineering
Ajou University, South Korea

Why we're all here!

Presentation Objectives

- Enjoy and Learn at Least One Thing!
- Discuss some modeling limitations that can inform safety research
- Motivate thoughts about safety modeling
- Provide background for following speakers

Overview

- Our practical problems (what we try to do)
  - Protect the public from undue harm
  - Understand *how* to efficiently improve safety (benefit to cost)
  - Understand *where* to improve safety

- Our modeling problems (in general)
  - Crashes are highly complex events
  - We have imperfect information
  - We have missing and superfluous information
  - We lack well grounded theories relating crashes to causes
  - We can conduct few (if any) controlled experiments

- Some methods and their limitations
  - Fundamentals
  - Before-after methods
  - Linear regression
  - Negative binomial/Poisson regression
  - Zero inflated models
  - Systems of equations
  - Neural nets/AI models

Our Practical Problems

- Protect the public from undue harm
- Understand *how* to efficiently improve safety (benefit to cost)
- Understand *where* to improve safety

Protect the public from undue harm:

This is worthy of an entire presentation; however, it implies:

1. ‘We’ must apply due diligence in providing for safe travel
2. ‘We’ must not neglect high risk situations
3. ‘We’ must not show preferential treatment

Understand *how* to efficiently improve safety (benefit to cost)

Which countermeasures are effective and when?

1. Policies
   - (Licensing, vehicle restrictions, speed zones, etc.)
2. Programs
   - (driver training, enforcement, driver check points, etc.)
3. Human Factors
   - (Vehicle design, information processing, illumination, etc)
4. Operations
   - (geometry, access management, signal phasing, etc.)
5. Roadside/infrastructure
   - (side slopes, objects, clear zones, materials, etc.)

Understand *where* to improve safety

1. Which situations have highest *crash risk*?
2. Which situations have highest *crash severity*?
3. Which situations have highest *improvement potential*?

(Note: '*situation*' replaces '*location*', e.g. a drunk driver is a situation)

Our Modeling Problems

- Crashes are highly complex events
- We possess imperfect information
- Missing and superfluous information plagues us
- Well articulated theories relating crashes to causes are lacking
- We can conduct few (if any) controlled experiments

Crashes are highly complex events

- Crashes are *outcomes* of a highly complex *process*.
- We rarely witness the *process*, and less rarely the *outcome*.
- We rely on someone's interpretation (and others' interpretations) and recording of the *process* and *outcome*.
- Critical elements in the *process* are highly interrelated.

We posses imperfect information

- Much data are imprecise, aggregated, or incorrect
- Many PDO crashes are unreported
- Some reporting details are inaccurate
- Volume/exposure data are imprecise
- Location information is notoriously inaccurate
- Travel speeds are approximate

Missing and superfluous information plagues us

- Human performance missing (vision, reaction times, distractions, mood, etc.)
- Potential contributing factors often missing (distractions, other motorists, sunset/sunrise, debris, etc.)
- BAC content often missing
- Drug role often missing

We lack well articulated theories relating crashes to causes

- Much of our knowledge comes from empirical work—which contains inconsistencies and suffers from data and analysis problems
- Because crashes are linked to humans theories are 'social' in nature (i.e. generally applied and full of exceptions)

We lack well articulated theories relating crashes to causes (cntd.)

- Lack of theories leads to much data mining (equivalent to searching for a piece of hay in a haystack)
- Analysis often gives rise to theories (reverse to scientific method)
- Lack of well articulated theories leads to difficulty in applying the scientific method

We conduct few (if any) controlled experiments

- Many factors that are known to affect safety cannot be controlled or manipulated:
  - Traffic volumes
  - Road users
  - Weather
  - Travel speeds
- These variables are some of the most influential factors on safety!

We conduct few (if any) controlled experiments (cntd.)

- Powerful experimental tools cannot be applied:
  - Random selection of experimental units
  - Random assignment of treatment groups
  - Elimination of extraneous variables
  - Manipulation of extraneous variables
- Result: Studies are plagued with:
  - Correlated variables
  - Self-selection of treatment groups
  - Non-random samples
  - Intrusion of effects of omitted variables

So, with all of these challenges (i.e. job security), what do we do?

We should:

- Understand the (modeling) limitations
- Report the limitations
- Continue to improve methods
- Continue consensus building
- Question conventional wisdom
- Commit to developing and rewarding safety expertise

Some Methods and Their Limitations

- Crash process fundamentals
- Before-after methods
- Linear regression
- Negative binomial/Poisson regression
- Zero inflated models
- Systems of equations
- Meta-analysis
- Neural nets/AI models

Crash Modeling Fundamentals

History and considerable research tells us that motor vehicle crashes are affected by:

- Driver inattentiveness and/or risk taking
- Traffic volumes (exposure to risk)
- Excessive speed (too fast for design/conditions)
- Roadway geometrics (lane widths, shoulders, medians, alignment, sight distance)
- Roadside obstacles/hazards
- Adverse weather conditions
- Signage/control/operations

1. Statistical models allow us to characterize the crash process
2. All statistical models are wrong, to some degree
3. **The difference between "grossly wrong" and "slightly wrong" models is not determined statistically**

Experience and Observation also tells us three very important things:

- Crash counts are stochastic
- The count of crashes in any given observation period reflects underlying safety AND random fluctuation
- Thus, the observed count in any given period is not a reliable indicator of the underlying safety of a site

Thus, to obtained expected safety in a stochastic process, we use:

Mean Crash Count = F (exposure, geometrics, speed, weather, etc.) + ERROR

Where,

ERROR is stochastic and includes the aggregate effect of omitted variables, measurement error, mis-specification error, and other possible effects

A stochastic model of crashes

- A crash is the result of a Bernoulli trial. Each time a vehicle enters an intersection (highway segment, ramp, etc.) it will either crash or not crash.
- A series of Bernoulli trials results in a binomial distribution (n “successes” of N trials):

$$P(Z = n) = \binom{N}{n} p^n (1 - p)^{N-n}$$

A better stochastic model of crashes

When N is large (e.g. AADT) and p is small (crash on any given trial), the Binomial distribution converges to the Poisson distribution:

$$P(Z = n) = \binom{N}{n}\left(\frac{\lambda}{N}\right)^{n}\left(1 - \frac{\lambda}{N}\right)^{N-p} \cong \frac{\lambda^{n}}{n!}e^{-\lambda}$$

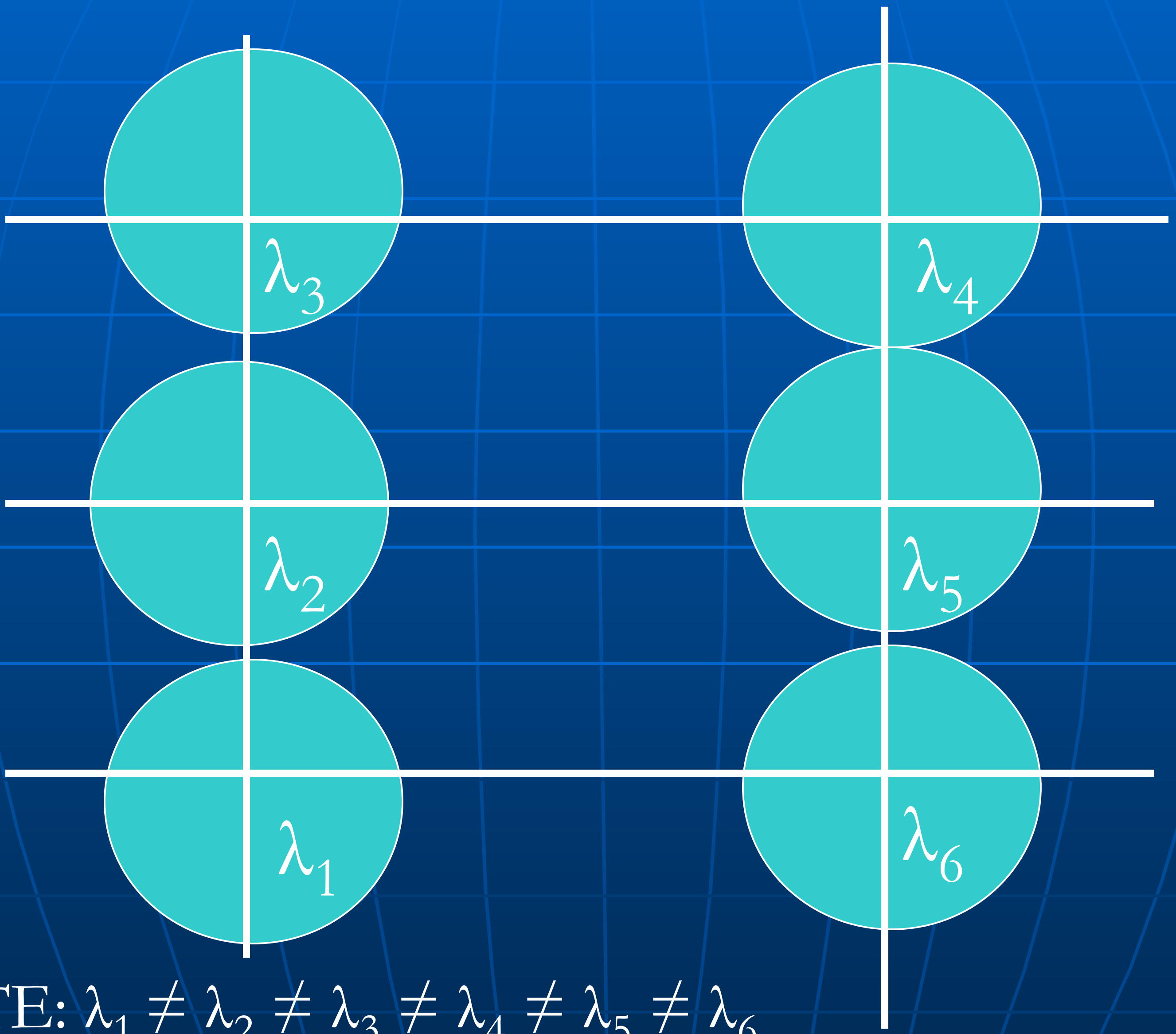
Expected Number of Crashes

- The Poisson distribution has one parameter, λ, which is the Poisson mean and variance.

- The requirement that the mean and variance are equal causes some problems:

- Consider a transportation network, where crashes occur as a Poisson process at various intersections, such that:

6 Intersections in Network



NOTE: $\lambda_1 \neq \lambda_2 \neq \lambda_3 \neq \lambda_4 \neq \lambda_5 \neq \lambda_6$

Why are the intersection mean crash rates different?

1. They vary as a function of observed variables (exposure, geometrics, speed, weather)

2. They vary due to the effects of UNOBSERVED variables on crashes (proximity to bars, police enforcement effects, driving population, vehicle fleet differences, etc.)

It is the variation in means across locations that leads to lack of fit when using the Poisson distribution.

The “best” stochastic model of crashes

A negative binomial distribution arises as a combination of gamma distributed heterogeneity of Poisson means.

This distribution is considered the current state of the practice for modeling over-dispersed crash data.

$$P\left(Y_k = n\right) = C\left(n-1, k-1\right) p^k \left(1-p\right)^{n-k}; \text{ for } n = k, k+1, k+2, \ldots$$

Some Methods and Their Limitations

- Linear regression
- Negative binomial/Poisson regression
- Zero inflated models
- Systems of equations
- Before-after methods
- Meta-analysis
- Neural nets/AI models

Methods Critique

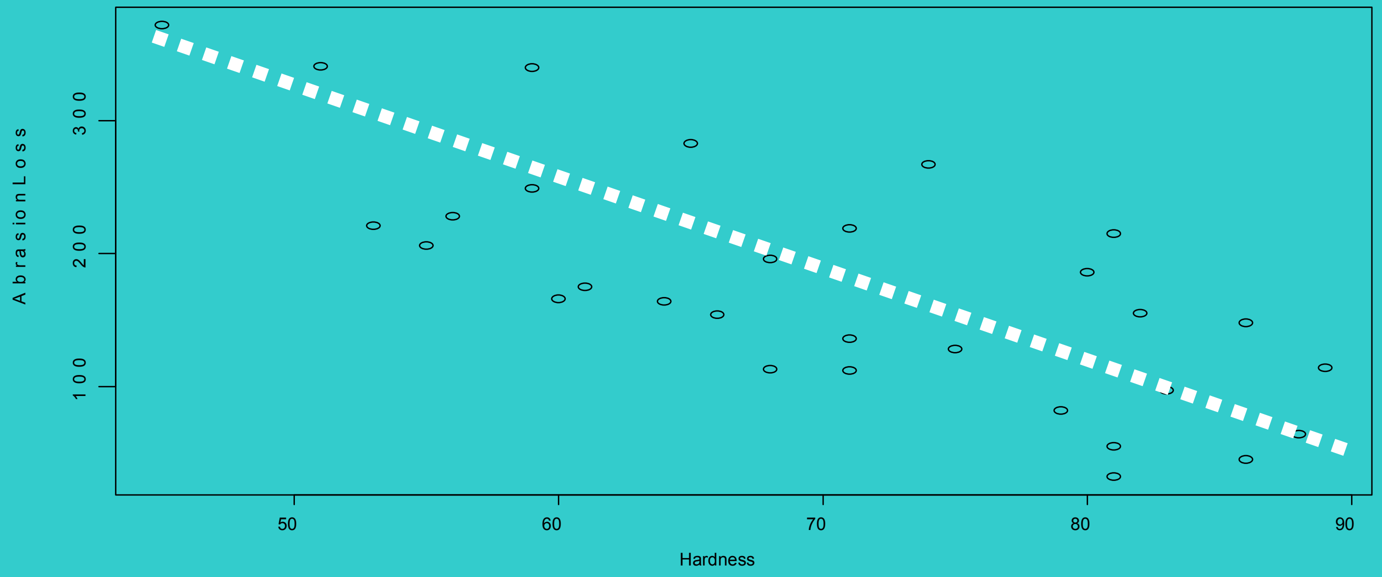
- Objectives/uses
- Strengths/Limitations
- Not methodology

Linear Regression

Objective:

To model a continuous variable as a function of covariates for explanation, prediction, or quality control.

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + \ldots + \beta_{P-1} X_{P-1,i} + \varepsilon_i$$



Linear Regression

Strengths

- Very flexible, can be used to model a variety of linear and non-linear relationships
- Assumptions fairly robust
- Multivariate
- Easy to learn (or is this a limitation?)
- Useful for illustrating the basics of modeling

Linear Regression

Weaknesses

- Suited for continuous variables only
- Can lead analyst astray
- Can show ‘good fit’ and be a ‘bad model’
- Limited in model forms (can’t handle both additive and multiplicative effects)
- Often abused (i.e. data mining)

Linear Regression: Leading the analyst astray

- Analysts are challenged to accommodate interactions, non-linear relationships, and non-additive behavior in models.
- Transformations are typically used to do this, however, transformations drastically change the fundamental character of assumed relationships in models—often without knowledge of the analyst

Leading the analyst astray: Example

Suppose the underlying function is thought to be:

$Y = \alpha \exp^{\beta X} + \varepsilon$ (which is non-linear)

a log transformation will give:

$$\begin{aligned}\ln(Y) &= \ln(\alpha \exp\beta X + \varepsilon) \\ &= \ln[(\alpha \exp\beta X)(1 + \varepsilon/\alpha \exp\beta X)] \\ &= \ln(\alpha) + \beta X + \ln(1+\varepsilon/\alpha \exp\beta X)).\end{aligned}$$

Although the model is linear, the error term is not the one specified in ordinary least squares regression because it is a function of X, α, and β, and is multiplicative.

Leading the analyst astray: Transformations

Result:

Transformations so often performed in linear regression (and other modeling platforms) can result in fundamental changes to the functional form of the model that regression assumptions are violated!

Regression Cautions

Goodness of fit statistics can be misleading, and may wrongly convince a modeler that the model is good (or bad).

T-statistics (and Z) give an indication of the magnitude of the association of a covariate relative to model ‘noise’: It does not indicate causality, proof, or finality.

Significance (by some subjective threshold) can indicate the presence of an omitted lurking variable; the effect of an extremely large dataset, an error in judgment, or a real effect.

Non-significance can indicate an error, a noisy dataset, collinearity with included variables, or lack of effect.

Good Regression Models

- Don’t violate regression assumptions
- Can be explained readily
- Have real-world interpretations
- Capture main effects and interactions
- Are selected based on theoretical appeal not ‘statistical fit’
- Are validated against other data
- Capture known effects

Negative Binomial/Poisson Regression

Objective:

Model count data as a function of covariates and capture over-dispersion in crash process by assuming gamma distributed means across locations/sites with Poisson crash processes.

Negative Binomial/Poisson Regression

Strengths:

- Gets around linear regression constraints
- Multiplicative terms
- Has been successful for modeling crash data (theoretical appeal, good fit)
- Requires some knowledge to estimate

Negative Binomial/Poisson Regression

Weaknesses:

- Can't accommodate additive effects
- Fewer feedback tools than linear reg.
- All the same weaknesses as linear regression: t-statistics, transformations, interactions, etc.

Zero-Inflated Models

Objective:

- Extension of Poisson and Negative Binomial: Model count process as dual-state process
- State 1: inherently safe (count = 0)
- State 2: Poisson/NB process (count = 0, 1, 2, 3, ……n)

Zero-Inflated Models

Strengths:

Can often fit crash data better and appears to account for “excess” zeroes often found in count data (i.e. zeroes not explained by Poisson or NB process)

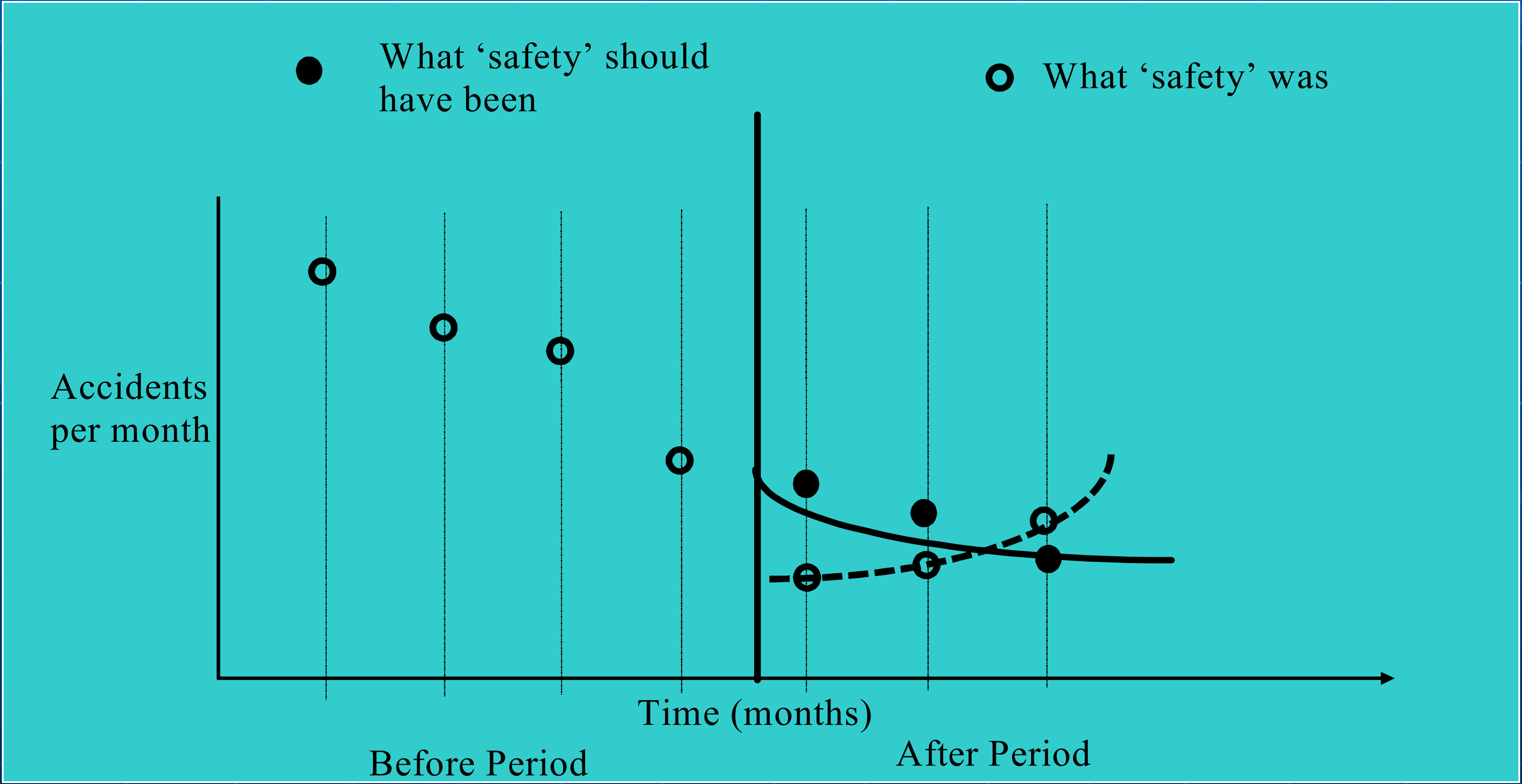
Zero-Inflated Models

Weaknesses:

- Hard to justify/support notion of 'inherently safe' sites (lacks theoretical appeal)
- Difficult to interpret and/or explain differences b/w dual states

Before-After Methods

Objective: Estimate the effect of a countermeasure, intervention, program, or policy on safety.



Before-After Methods

Strengths:

- Range of sophistication to suit a variety of analyses
- Produces most reliable estimate of effectiveness (state of the art)
- Method utilizes some aspects/strengths of experimental methods
- Can account for regression to the mean (self selection problem)
- To be discussed more later….

Before-After Methods

Limitations:

- Some countermeasures, interventions, programs, and policies difficult to evaluate in B-A framework
- Costly and time consuming
- A ‘one at a time’ approach
- Requires significant expertise (or is this a strength?)

Meta-Analysis

Objective:

To compile and analyze the results of many/multiple studies to estimate the composite effect of cumulative studies.

Uses things such as sample sizes, experimental ‘shortcomings’, variance estimates, effect sizes, etc. to derive a weighted average effect

Meta-Analysis

Strengths:

- Gets away from relying on a small number of studies
- Quantifies ‘consensus’
- Is rigorous and defensible
- Has sound methodology

Meta-Analysis

Weaknesses:

- Difficult to apply in practice: much desired information is not published and is unavailable (e.g. covariance matrices)
- Hard to assign ‘weights’ to things like selection bias, missing variables, non-representative samples, etc.
- Involves subjective assignment of weights which can become key points of contention

Neural nets/AI models

Objective:

Borrow from AI community methods for fitting multivariate data to response.

Relies on training sets and different type of algorithms to fit data.

Neural nets/AI models

Strengths:

- Can attain high prediction capabilities, often higher than ‘conventional’ models
- Fairly easy to estimate
- May be extremely useful for “simulation” experiments (e.g. simulating crashes on a network under a variety of conditions)

Neural nets/AI models

Weaknesses:

- Don’t provide insight into underlying *process* (which is often important)
- Can’t determine if model is just ‘over fit’ to data (like a regression models with all interaction terms)
- Lacks useful interpretation/insight for safety policy/countermeasure/program implementation

Conclusions: Limitations to Modeling

- All models have strengths and weaknesses
- We must know them!
- Statistical results don't determine the 'best' model
- Theory, engineering knowledge, and sound reasoning accompany good statistics and models

Conclusions: Limitations to Modeling (cntd.)

- Modeling requires the analyst to consider many things:
  - interactions
  - model functional form (relationships)
  - model assumptions
  - experimental shortcomings
  - Selection bias
  - interpretability
  - Included and omitted variables
  - Appropriate modeling methodology

Remember:

- Understand the (modeling) limitations
- Report the limitations
- Continue to improve methods
- Continue consensus building
- Question conventional wisdom
- Commit to developing and rewarding safety expertise